

# Clusters of sexual behaviour in HIV-positive men who have sex with men reveal highly dissimilar time trends

Luisa, Salazar-Vizcaya<sup>1\*</sup>; Katharina, Kusejko<sup>2,3</sup>; Axel J, Schmidt<sup>4,5</sup>; Germán, Carrillo-Montoya<sup>6</sup>; Dunja, Nicca<sup>7</sup>; Gilles, Wandeler<sup>1</sup>; Dominique L, Braun<sup>2,3</sup>; Jan, Fehr<sup>2</sup>; Katharine EA, Darling<sup>8</sup>; Enos, Bernasconi<sup>9</sup>; Patrick, Schmid<sup>4</sup>; Huldrych F, Günthard; Roger D, Kouyos and Andri, Rauch

1 Department of Infectious Diseases, Bern University Hospital Inselspital, University of Bern, Bern, Switzerland;

2 Division of Infectious Diseases and Hospital Epidemiology, University Hospital Zurich, Zurich, Switzerland;

3 Institute of Medical Virology, University of Zurich, Zurich, Switzerland;

4 Division of Infectious Diseases and Infection Control, Cantonal Hospital St. Gallen, Switzerland;

5 Sigma Research, London School of Hygiene and Tropical Medicine, United Kingdom;

6 Alpiq Energy AI, Olten, Solothurn, Switzerland;

7 Institute of Nursing Science, University of Basel, Basel, Switzerland;

8 Infectious Diseases Service, Department of Medicine, University Hospital of Lausanne (CHUV), Switzerland;

9 Division of Infectious Diseases, Lugano Regional Hospital, Lugano, Switzerland;

© The Author(s) 2019. Published by Oxford University Press for the Infectious Diseases Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

**¶Corresponding author:**

Luisa Salazar-Vizcaya, PhD

Tel: + 41 31 632 7738

Bern University Hospital, Inselspital

Universitätsklinik für Infektiologie, Personalhaus 6

3010 Bern, Switzerland

Email: [luisapaola.salazarvizcaya@insel.ch](mailto:luisapaola.salazarvizcaya@insel.ch)

**Summary:** We propose a method that uses machine learning to identify behavioural groups in men who have sex with men based on the recognition of condom use patterns over time. It led to the identification of clusters with dissimilar behavioural trends.

## Abstract

**Background:** Separately addressing specific groups of people who share patterns of behavioural change might increase the impact of behavioural interventions to prevent transmission of sexually transmitted infections. We propose a method based on machine learning to assist the identification of such groups among men who have sex with men (MSM).

**Methods:** By means of unsupervised learning, we inferred “behavioural clusters” based on the recognition of similarities and differences in longitudinal patterns of condomless anal intercourse with non-steady partners (nsCAI) in the Swiss HIV cohort study over the last 18 years. We then used supervised learning to investigate whether sociodemographic variables could predict cluster membership.

**Results:** We identified four behavioural clusters. The largest behavioural cluster (Cluster 1) contained 53% of the study population and displayed the most stable behaviour. Cluster 3 (17% of the study population) displayed consistently increasing nsCAI. Sociodemographic variables were predictive for both of these clusters. The other two clusters displayed more drastic changes: nsCAI frequency in Cluster 2 (20% of the study population) was initially similar to that in Cluster 3, but accelerated in 2010. Cluster 4 (10% of the study population) had significantly lower estimates of nsCAI than all other clusters until 2017, when it increased drastically, reaching 85% by the end of the study period.

**Conclusions:** We identified highly dissimilar behavioural patterns across behavioural clusters, including drastic, atypical changes. These patterns suggest that the overall

increase in the frequency of nsCAI is largely attributable to two clusters, accounting for a third of the population.

**Key words:** sexual behaviour, clusters, STI, men who have sex with men, HIV, condom

## Introduction

Sexual behaviour among men who have sex with men (MSM) is heterogeneous and dynamic [1-3]. Recent changes in sexual behaviour have been associated with increases in sexually transmitted infections (STI) among MSM [4-8]. Yet, potentially oversimplified assumptions on exposure to STI transmission are common [9-11]. Lack of understanding and misconceptions about choices regarding sexual encounters are likely to have limited the efficiency and efficacy of public health interventions, as well as the accuracy of mathematical modelling projections, because they may have failed to capture underlying drivers of risk-taking and heterogeneity.

Separately addressing specific groups of people who share similar attitudes towards risk-taking might increase the impact of interventions aimed at preventing exposure to STI, and reducing their transmission. Recognizing such attitudes in individuals may be critical to design interventions that are more effective [12, 13]. For instance, previous modelling of individual decision-making considering game theory found that emerging behaviours in populations are highly sensitive to the value/worth that individuals attribute to condomless sex with HIV-serodiscordant partners [14]. Yet, assumptions regarding risk-taking have often been based on individuals' sociodemographic characteristics [15-17]. *A-priori* categorizations are subject to stereotypical representations of people. While these approaches sometimes succeed in identifying predictors of sexual behaviour, it is less clear to what extent such characteristics drive decisions regarding sexual encounters [18, 19], their role in the conformation of beliefs systems, knowledge, motivation and behavioural change [20]. Alternative approaches to questioning routinely collected data could help identify and contextualize otherwise unnoticed heterogeneity in behavioural change.

As opposed to conventional approaches, our work exploits an intuitive principle. We used unsupervised learning to infer “behavioural clusters” based purely on the recognition of longitudinal patterns of Condomless Anal Intercourse with non-steady partners (nsCAI) over the last 18 years in the Swiss HIV Cohort Study (SHCS; [www.shcs.com](http://www.shcs.com)) [21]. We then characterized the relationship between cluster memberships and sociodemographic characteristics by means of supervised machine learning methods.

## Methods

**Systematic data collection on sexual behaviour in the SHCS:** The Swiss HCV Cohort Study (SHCS, [www.shcs.ch](http://www.shcs.ch)) is a nationwide prospective cohort that routinely collects behavioural, laboratory and clinical data from HIV-positive persons aged  $\geq 16$  years since 1988. Individual data are recorded at study entry and every 6 months thereafter. We estimate that more than 80% of all MSM currently diagnosed with HIV in Switzerland are followed in the cohort [8, 22].

The SHCS has been approved by the ethics committee of the participating institutions (Kantonale Ethikkommission Bern, Ethikkommission des Kantons St. Gallen, Comité départemental d'éthique des spécialités médicales et de médecine communautaire et de premier recours, Hôpitaux Cantonale de Genève, Kantonale Ethikkommission Zürich, Repubblica e Cantone Ticino—Comitato Etico Cantonale, Commission cantonale d'éthique de la recherche sur l'être humain, Canton de Vaud, Lausanne, Ethikkommission beider Basel for the SHCS and Kantonale Ethikkommission Zürich for the ZPHI). Written informed consent had been obtained from all participants.

**Behavioural matrix:** This matrix contains trajectories of condomless anal intercourse with non-steady partners (nsCAI) over time for each patient. A binary variable determined the “nsCAI state” of each patient at time of interviewing (i.e. engaging in nsCAI: yes or not). The succession of such states defined a patient’s nsCAI trajectory. The state was determined according to the patient’s answers to the questions: i) Did you have sex with occasional partners in the last six months?; if yes, ii) Did you have anal intercourse with these partners?; if yes, iii) Did you use condoms all the time?. When changes in the nsCAI state were recorded, it was flipped at the midpoint between the two discrepant registries. This status only changed based on new information i.e., a patient who had been determined to engage in nsCAI was assumed to do so until the opposite could be determined based on a new interview. At least one report of nsCAI, a minimum of two nsCAI assessment records, and a follow-up time of 2 years or more were required for inclusion in these analyses.

Individual nsCAI trajectories were combined into a standardized matrix (behavioural matrix), which we constructed by transforming them using piecewise functions. The behavioural matrix recorded binary patients’ statuses over standardized 3-monthly-updated intervals with a fixed time span (years 2000 to 2018.25). Consequently, the behavioural matrix was sized (# of persons X # of time-periods). The matrix only had valid entries for the periods where the patient in question was under follow-up and the outcome of the nsCAI assessment conclusive. Figure 3 exemplifies nsCAI trajectories for 12 randomly selected patients.

**Inference of Behavioural Clusters:** Based on the notion that similar trajectories of nsCAI may indicate concordant behavioural patterns, we inferred clusters of nsCAI trajectories (as recorded in the status matrix) by means of agglomerative hierarchical clustering. We used a

binary metric (i.e. proportion of concordant bits, also known as Jaccard distance) for computing the distance matrix, and the Ward's method as agglomeration criterion [23].

To increase the likelihood of successful agglomeration in the status matrix, we considered information from mid-2001 (as sexual behaviour questionnaires were first introduced in 2000). We utilized the R function *hclust* [24] to produce the clusters presented in this manuscript.

**Discriminatory power of sociodemographic characteristics:** We investigated whether a set of sociodemographic variables available from the cohort data could predict to which behavioural cluster a patient would belong. We simultaneously used Boosted Decision Trees (BDT), k-Nearest Neighbours (kNN) and Maximum Likelihood Methods (ML) to seek for associations between cluster membership and a set of sociodemographic variables including age, (calendar) time of registration, origin and education level. We summarized the aggregated outcome of these analyses with Receiver Operating Characteristic (ROC) curves. The Toolkit for Multivariate Data Analysis with ROOT (TMVA; <https://root.cern.ch/http://tmva.sourceforge.net/>) served as the main tool for these analyses. In analogy with the “particle discovery” problem, we refer to background rejection and signal efficiency in this context as successful rejection of persons not belonging to a behavioural cluster, and successful classification of persons belonging to a behavioural cluster, respectively.

Finally, we implemented a toy Monte Carlo “permutations test” for the significance (measured as a p-value) of the classification based on sociodemographic variables. This test consisted of the iterative assigning of cluster labels to individuals at random, while conserving clusters' size and computing the corresponding (simulated) ROC curves. We did this 3,000 times for each cluster label and estimated the resulting p-values as the number of



iterations in which the simulated area under the ROC curve (AUC ROC) equalled or exceeded that of the original classification.

Computations were implemented in R (version 3.4.4), TMVA (version 4.2, running under Root version: 6.14) and python (version 3.7, libraries: pandas[25], numpy[26], scipy [27]). All software ran under Arch Linux x86\_64 4.18.1.

## Results

Of 6025 HIV-diagnosed MSM ever active in the SHCS between 2000 and May 2018, 2766 reported nsCAI at least once. Of those, 2539 had at least 2 years of follow-up and two or more nsCAI data records. The analyses in this manuscript were based on the latter group. Patients' median age at enrolment in the cohort was 36 years (interquartile range [IQR]: 30-43), median follow-up over the study period was 10 years (IQR: 6-16 years), and 44% (1105/2539) reported (ever) using of recreational drugs. **Table 1** shows characteristics of the patients included in the analyses.

**Behavioural matrix and aggregated trends (Figure 1):** Figure 1A displays individual nsCAI trajectories. The median distance between trajectories defined by a binary metric was 0.8 (IQR: 0.4-1.0). The overall fraction of patients with nsCAI increased from 20% (95% confidence interval [CI]:15-24%) in 2001 to 67% (95% CI:64-70%) in 2018 (Figure 1B).

**Distinct behavioural clusters (Figure 2):** We assessed nsCAI trajectories for the top four hierarchies' clusters. In this case, this was equivalent to a cut at 1/3 of the full dendrogram height (Figure S1). Figure 3 shows exemplary trajectories of nsCAI randomly retrieved from each of the resulting clusters.

We found dissimilar behaviours across clusters (coloured in Figures 2A and 2B). All clusters displayed increasing nsCAI over the study period, and Cluster 2 (blue) was the only one without a net increase in nsCAI between 2014 and 2018 (Figure 2B). Until about 2006, nsCAI trends for all clusters overlapped (Figure 2B). Cluster 4 (red) was the first to distinguish itself from the others, with consistently lower nsCAI until 2016 (Figure 2B). Median pairwise distances were: 6.0 (IQR:5.1-7.1), 4.9 (IQR:3.2 - 5.6), 2.9 (IQR:1.3 - 3.5), 5.4 (IQR:2.3 - 5.4) for clusters 1 to 4 respectively.

In the largest cluster (Cluster 1, violet, 53% of total number of included patients) the fraction with nsCAI varied the least, going from 25% (95% CI: 19-31%) in 2001 to 43% (95% CI:39-47%) in 2018 (Figure 2B). Two clusters displayed consistently higher and increasing fractions (Cluster 2, blue and Cluster 3, green; accounting for 20% and 17% of total number of included patients, respectively). The nsCAI curves for these two clusters were similar at the beginning, but they seem to split in 2010: nsCAI grew faster in Cluster 2 than it did in Cluster 3. Interestingly while nsCAI declined from 2014 onwards in cluster 2, it continued to increase in cluster 3. Both nsCAI curves appeared to stabilize in 2016. In Cluster 4 (10% of total number of included patients), nsCAI remained below 13% until 2017, when it rose sharply to reach 85% (95% CI:80-90%) in 2018.

#### **Discriminatory power of traditional sociodemographic characteristics (Figures 4-5):**

These analyses, which were independently performed for each cluster in Figure 2, suggests that sociodemographic characteristics had discriminatory power for Clusters 1 and 3 (70% of the total number of patients) but not for Clusters 2 and 4, as evidenced by the ROC curves in Figure 4 and by the toy Monte Carlo permutations tests in Figure 5. This test assessed the significance of the classification based on sociodemographic variables. P-values estimated through the permutation test ranged between 0.001 and 0.73 for Cluster 2, and between 0.001 and 0.4 for cluster 4. This suggests no association between cluster membership and

sociodemographic variables in these two clusters. In the same test for Clusters 1 and 3, all simulated AUC were below those of the original classification. This constitutes the strongest association we found between cluster membership and sociodemographic variables.

In order to identify potential prominent features, and in addition to the above described use of three machine learning algorithms, we directly examined the distributions of sociodemographic variables across clusters and their correlation, which are shown in Figure 6. Note that in panel F, the distribution of year of registration in the cohort has separated peaks for Clusters 1 and 3. Members of Cluster 3 (green) had the latest average registration date across clusters, which suggests more recent HIV infections. By contrast, members of Cluster 1 (violet) had the earliest average registration date across clusters. As depicted above, Clusters 1 and 3 (violet and green) also yielded the strongest discriminatory power of sociodemographic variables. Neither visual inspection nor statistical tests indicated further prominent differences between clusters regarding variable distribution/correlation.

## Discussion

Behavioural clustering purely based on individual trajectories of Condomless Anal Intercourse with non-steady partners (nsCAI) in MSM suggest that the continuous overall increase in nsCAI observed over the last 18 years was the consequence of collective, yet heterogeneous behaviours. These included drastic changes occurring over time-periods that differed across clusters (Figure 2). The overall increase in this practice over the last 18 years is largely attributable to two behavioural clusters accounting for a third of the population (Clusters 2 and 3). The largest behavioural cluster contained 53% of the study

population and displayed the most stable behaviour over time. Sociodemographic variables were predictive of cluster membership for two behavioural clusters containing 70% of the study population, but not for the remaining two clusters, which displayed the most drastic changes over time.

To the best of our knowledge, this is the first published study to infer risk groups and to depict behavioural trends based purely on trajectories of sexual practices (nsCAI in this case). It did not assume *a-priori* that persons' characteristics such as age, origin, level or education or year of HIV diagnosis explained their choices regarding sex. The availability of longitudinal, long-term records of nsCAI, which the SHCS has collected for almost 18 years, enabled this analysis.

The method outlined here is intuitive. Grounding the analyses on nsCAI alone allowed a compact presentation of the method and facilitated results interpretation. While considering only one dynamic variable (nsCAI) may be seen as a limitation, this variable, often recorded in studies on sexual behaviour, has shown to be a powerful predictor of other sexual behaviours and of STI transmission [4, 5, 7, 28, 29]. Adapting this method to include more variables is straightforward, and it is suitable for any setting with available longitudinal data on nsCAI or other quantities including clinical outcomes. We evaluated for discriminatory power variables available from the SHCS data, which are most commonly recorded in other longitudinal studies. However, we cannot exclude the existence of more predictive, yet unmeasured sociodemographic characteristics. Moreover, this approach can be used within an explanatory mixed method research design [30, 31]: Hypotheses generated based on clusters could be explained by qualitative data that provide more comprehensive insights into behavioural change. Suitable methods alternative to our algorithmic approach (*i.e.*, hierarchical clustering) include model-based clustering such as

longitudinal latent class analyses (LLCA). However, the outcome of hierarchical clustering offers an in-depth view of the risk structure of the population, and unlike LLCA, algorithmic clustering does not require model fitting, or hypothesis regarding data structure that could constrain the classification.

This study does not deal with the attribution of specific underlying mechanisms that bound patients within clusters together. A possible explanation for the behavioural patterns depicted in this manuscript is that patients sharing behavioural clusters respond and adapt similarly to external information such as messages from the media, public campaigns aimed at reducing exposure to STI transmission, health care provider information and scientific releases. For example, the acceleration in nsCAI in Cluster 2 (Figure 2B, blue) coincided with the diffusion of the Swiss statement (part of a publication by Swiss researchers which stated that people with HIV were not infectious if they were on effective antiretroviral therapy for at least six months and without any STI) [32]. This concept is closely related to the U(undetectable) = U(untransmissible) message [33, 34], which has been widely supported by subsequent studies [35]. The sharp rise of nsCAI in Cluster 4 in 2017 may be associated with awareness resulting from the publication of landmark studies confirming the efficacy of PrEP (from 2015 onwards [36-38]) [39, 40] and with the rapid spread of chemsex (sexual activity under the influence of stimulant drugs such as methamphetamine, mephedrone, GHB/GBL or Ketamine) in Switzerland [41]. Importantly, decreasing condom use following the rollout of PrEP has been documented [40]. We believe the method outlined in this manuscript could help identify triggers of behavioural change. Of note, the remarks in this paragraph are of a hypothetical nature and proving or disproving them is beyond the scope of this manuscript. A further study aimed at assessing these hypotheses by means of qualitative research is warranted.

Finally, sexually transmitted infections propagate along sexual networks [42]. But sexual behaviour may change unevenly within sexual networks if individuals sharing a sexual network do not share decision-making mechanisms and sexual behaviours [43]. We therefore think that the method depicted in this manuscript is complementary to those concerning the characterization of contact structures (e.g. inferred from transmission networks) [44-46].

In summary, we identified behavioural clusters based purely on the recognition of similarities and differences in longitudinal patterns of change in condomless anal intercourse with non-steady partners. The method we proposed could help identify key target populations for behavioural interventions and meaningful risk groups for modelling of sexually transmitted infections. Both are key to achieving optimal allocation of resources to fight STI transmission. Available sociodemographic variables were found to be good predictors of behavioural clustering for the majority of the population, but not for those men who displayed the most drastic changes in sexual behaviour over time. A complete identification of such risk groups will require characterizing patients within clusters. For that purpose, and to understand drivers of changes in sexual behaviour within clusters, further mixed methods studies combining quantitative and qualitative research are warranted.

**Author contribution:** LSV and AR designed the study. LSV and GCM performed the analyses. LSV and AR prepared the first draft manuscript, which was revised by all co-authors. GW, DLB, JF, KEAD, EB, PS, HFG and AR contributed to data acquisition. All authors contributed to the interpretation of analyses outcomes.

**Acknowledgements:** We thank all patients, physicians and nurses associated with the Swiss HIV Cohort Study. Members of the Swiss HIV Cohort Study: Anagnostopoulos A, Battegay M, Bernasconi E, Böni J, Braun DL, Bucher HC, Calmy A, Cavassini M, Ciuffi A, Dollenmaier G, Egger M, Elzi L, Fehr J, Fellay J, Furrer H (Chairman of the Clinical and Laboratory Committee), Fux CA, Günthard HF (President of the SHCS), Haerry D (deputy of “Positive Council”), Hasse B, Hirsch HH, Hoffmann M, Hösli I, Huber M, Kahlert C, Kaiser L, Keiser O, Klimkait T, Kouyos RD, Kovari H, Ledergerber B, Martinetti G, Martinez de Tejada B, Marzolini C, Metzner KJ, Müller N, Nicca D, Paioni P, Pantaleo G, Perreau M, Rauch A (Chairman of the Scientific Board), Rudin C (Chairman of the Mother & Child Substudy), Scherrer AU (Head of Data Centre), Schmid P, Speck R, Stöckle M, Tarr P, Trkola A, Vernazza P, Wandeler G, Weber R, Yerly S.

**Disclaimer:** The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Funding:** This work was performed within the framework of the SHCS, supported by the Swiss National Science Foundation (SNF grants #33CSC0-108787 and #324730\_179567) and by the SHCS grant #823. DLB was supported by the University of Zurich’s Clinical Research

Priority Program Viral Infectious Diseases. RDK was supported by the SNSF (grant # BSSGI0\_155851).



**Competing interests:** HFG has received unrestricted research grants from Gilead Sciences, Roche; a grant from Systems X; grants from SHCS research foundation, NIH, Yvonne Jacob Foundation, personal fees from DSMB / Consulting for Merck, Gilead, Teva, Sandoz, outside the submitted work. AR reports support for advisory boards and/or travel grants from Janssen-Cilag, MSD, Gilead Sciences, Abbvie, and Bristol-Myers Squibb, and an unrestricted research grant from Gilead Sciences. All remuneration went to his home institution and not to AR personally. DLB reports support for advisory boards and/or travel grants from MSD, ViiV, and Gilead Sciences. RDK reports personal fees from Gilead Sciences. JF is a member of the Federal Commission for Sexual Health and reports support for advisory boards and/or travel grants from Janssen-Cilag, MSD, Gilead Sciences, Abbvie, Merck, ViiV Healthcare and Bristol-Myers Squibb, and an unrestricted research grant from Gilead Sciences. All remuneration went to his home institution and not to JF personally. GW reports grants from Gilead Sciences and AbbVie. AJS reports part time employment in a governmental agency (Swiss Federal Office of Public Health). KEAD reports contributions to her home institution from Gilead Sciences, grants from Merck Sharp and Dohme. EB reports payments to his home institution from Gilead Sciences, MSD, ViiV Healthcare, Pfizer, Sandoz and grants from Gilead Sciences, MSD, ViiV and Pfizer.

All these relationships were outside this work.

## References

1. Nostlinger C, Platteau T, Bogner J, et al. Implementation and Operational Research: Computer-Assisted Intervention for Safer Sex in HIV-Positive Men Having Sex With Men: Findings of a European Randomized Multi-Center Trial. *J Acquir Immune Defic Syndr* **2016**; 71(3): e63-72.
2. Sullivan PS, Carballo-Diequez A, Coates T, et al. Successes and challenges of HIV prevention in men who have sex with men. *Lancet* **2012**; 380(9839): 388-99.
3. Nostlinger C, Niderost S, Platteau T, et al. Sexual Protection Behavior in HIV-Positive Gay Men: Testing a Modified Information-Motivation-Behavioral Skills Model. *Arch Sex Behav* **2011**; 40(4): 817-27.
4. Rauch A, Rickenbach M, Weber R, et al. Unsafe sex and increased incidence of hepatitis C virus infection among HIV-infected men who have sex with men: the Swiss HIV Cohort Study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2005**; 41(3): 395-402.
5. Wandeler G, Gsponer T, Bregenzer A, et al. Hepatitis C virus infections in the Swiss HIV Cohort Study: a rapidly evolving epidemic. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2012**; 55(10): 1408-16.
6. Sullivan PS, Hamouda O, Delpech V, et al. Reemergence of the HIV epidemic among men who have sex with men in North America, Western Europe, and Australia, 1996-2005. *Ann Epidemiol* **2009**; 19(6): 423-31.
7. Shilaih M, Marzel A, Braun DL, et al. Factors associated with syphilis incidence in the HIV-infected in the era of highly active antiretrovirals. *Medicine (Baltimore)* **2017**; 96(2): e5849.
8. van Sighem A, Vidondo B, Glass TR, et al. Resurgence of HIV Infection among Men Who Have Sex with Men in Switzerland: Mathematical Modelling Study. *PLoS One* **2012**; 7(9): 10.

9. Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. *Lancet* **2009**; 373(9657): 48-57.
10. Bezemer D, de Wolf F, Boerlijst MC, et al. A resurgent HIV-1 epidemic among men who have sex with men in the era of potent antiretroviral therapy. *Aids* **2008**; 22(9): 1071-7.
11. Salazar-Vizcaya L, Kouyos RD, Zahnd C, et al. Hepatitis C Virus Transmission Among Human Immunodeficiency Virus-Infected Men Who Have Sex With Men: Modeling the Effect of Behavioral and Treatment Interventions. *Hepatology* **2016**; 64(6): 1856-69.
12. Ferguson N. Capturing human behaviour. *Nature* **2007**; 446(7137): 733.
13. Henry CJ, Koopman JS. Strong influence of behavioral dynamics on the ability of testing and treating HIV to stop transmission. *Sci Rep* **2015**; 5: 9467.
14. Tully S, Cojocaru M, Bauch CT. Sexual behavior, risk perception, and HIV transmission can respond to HIV antiviral drugs and vaccines through multiple pathways. *Sci Rep* **2015**; 5: 15411.
15. Avila D, Keiser O, Egger M, et al. Social Meets Molecular: Combining Phylogenetic and Latent Class Analyses to Understand HIV-1 Transmission in Switzerland. *Am J Epidemiol* **2014**; 179(12): 1514-25.
16. Golub SA, Kowalczyk W, Weinberger CL, Parsons JT. Preexposure prophylaxis and predicted condom use among high-risk men who have sex with men. *J Acquir Immune Defic Syndr* **2010**; 54(5): 548-55.
17. Gregson S, Garnett GP, Nyamukapa CA, et al. HIV decline associated with behavior change in eastern Zimbabwe. *Science* **2006**; 311(5761): 664-6.
18. Millett GA, Flores SA, Peterson JL, Bakeman R. Explaining disparities in HIV infection among black and white men who have sex with men: a meta-analysis of HIV risk behaviors. *Aids* **2007**; 21(15): 2083-91.
19. Valleroy LA, MacKellar DA, Karon JM, et al. HIV prevalence and associated risks in young men who have sex with men. Young Men's Survey Study Group. *JAMA* **2000**; 284(2): 198-204.

20. Stolte IG, Dukers NH, Geskus RB, Coutinho RA, de Wit JB. Homosexual men change to risky sex when perceiving less threat of HIV/AIDS since availability of highly active antiretroviral therapy: a longitudinal study. *Aids* **2004**; 18(2): 303-9.
21. Kouyos RD, Hasse B, Calmy A, et al. Increases in Condomless Sex in the Swiss HIV Cohort Study. *Open Forum Infect Dis* **2015**; 2(2): 4.
22. Kohler P, Schmidt AJ, Cavassini M, et al. The HIV care cascade in Switzerland: reaching the UNAIDS/WHO targets for patients diagnosed with HIV. *Aids* **2015**; 29(18): 2509-15.
23. Romesburg C. Cluster analysis for researchers: Lulu. com, **2004**.
24. Murtagh F, Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification* **2014**; 31(3): 274-95.
25. McKinney, Wes. "Data structures for statistical computing in python." *Proceedings of the 9th Python in Science Conference*. Vol. 445. 2010.
26. Travis E, Oliphant. *A guide to NumPy*, USA: Trelgol Publishing, (2006).
27. Jones E, Oliphant E, Peterson P, et al. *SciPy: Open Source Scientific Tools for Python*, 2001-, <http://www.scipy.org/> [Online; accessed 2018-11-01].
28. Traeger MW, Schroeder SE, Wright EJ, et al. Effects of Pre-exposure Prophylaxis for the Prevention of Human Immunodeficiency Virus Infection on Sexual Risk Behavior in Men Who Have Sex With Men: A Systematic Review and Meta-analysis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **2018**; 67(5): 676-86.
29. Pisani E, Girault P, Gultom M, et al. HIV, syphilis infection, and sexual practices among transgenders, male sex workers, and other men who have sex with men in Jakarta, Indonesia. *Sex Transm Infect* **2004**; 80(6): 536-40.
30. Research design: qualitative, quantitative and mixed methods approaches Research design: qualitative, quantitative and mixed methods approaches Creswell John W Sage 320 pound29 0761924426 0761924426 [Formula: see text]. *Nurse Res* **2004**; 12(1): 82-3.

31. Ruark A, Fielding-Miller R. Using Qualitative Methods to Validate and Contextualize Quantitative Findings: A Case Study of Research on Sexual Behavior and Gender-Based Violence Among Young Swazi Women. *Glob Health Sci Pract* **2016**; 4(3): 373-83.
32. Vernazza P, Hirschel B, Bernasconi E, Flepp M. Les personnes séropositives ne souffrant d'aucune autre MST et suivant un traitement antirétroviral efficace ne transmettent pas le VIH par voie sexuelle. *Bull Med Suisse* **2008**; 89: 165-9.
33. <https://www.niaid.nih.gov/news-events/undetectable-equals-untransmittable>.
34. Cohen MS, Chen YQ, McCauley M, et al. Antiretroviral Therapy for the Prevention of HIV-1 Transmission. *N Engl J Med* **2016**; 375(9): 830-9.
35. Rodger AJ, Cambiano V, Bruun T, et al. Sexual Activity Without Condoms and Risk of HIV Transmission in Serodifferent Couples When the HIV-Positive Partner Is Using Suppressive Antiretroviral Therapy. *JAMA-J Am Med Assoc* **2016**; 316(2): 171-81.
36. Molina JM, Charreau I, Spire B, et al. Efficacy, safety, and effect on sexual behaviour of on-demand pre-exposure prophylaxis for HIV in men who have sex with men: an observational cohort study. *Lancet HIV* **2017**; 4(9): e402-e10.
37. Sagaon-Teyssier L, Suzan-Monti M, Demoulin B, et al. Uptake of PrEP and condom and sexual risk behavior among MSM during the ANRS IPERGAY trial. *AIDS care* **2016**; 28 Suppl 1: 48-55.
38. McCormack S, Dunn DT, Desai M, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *Lancet* **2016**; 387(10013): 53-60.
39. Hampel B, Kusejko K, Braun DL, Harrison-Quintana J, Kouyos R, Fehr J. Assessing the need for a pre-exposure prophylaxis programme using the social media app Grindr(R). *HIV Med* **2017**; 18(10): 772-6.
40. Holt M, Lea T, Mao L, et al. Community-level changes in condom use and uptake of HIV pre-exposure prophylaxis by gay and bisexual men in Melbourne and Sydney, Australia: results of repeated behavioural surveillance in 2013-17. *Lancet HIV* **2018**; 5(8): e448-e56.

41. Hampel, B; Kusejko, K; Kouyos, R; Boeni, J; Flepp, M; Stoeckle, M; Conen, A; Béguelin, C; Künzler-Heule, P; Nicca, D; Schmidt, A; Nguyen, H; Delaloye, J; Rougemont, M; Bernasconi, E; Rauch, A; Günthard, H; Braun, D; Fehr, J (Zurich, Switzerland); Chemsex drugs on the rise among MSM: a longitudinal analysis of the Swiss HIV cohort study from 2007 to 2017; HIV drug therapy Glasgow 2018 Poster number P076.
42. Kouyos RD, Rauch A, Boni J, et al. Clustering of HCV coinfections on HIV phylogeny indicates domestic and sexual transmission of HCV. *Int J Epidemiol* **2014**; 43(3): 887-96.
43. Liljeros F, Edling CR, Amaral LA, Stanley HE, Aberg Y. The web of human sexual contacts. *Nature* **2001**; 411(6840): 907-8.
44. Bai W-J, Zhou T, Wang B-H. Interplay between HIV/AIDS epidemics and demographic structures based on sexual contact networks. *International Journal of Modern Physics C* **2007**; 18(06): 1025-45.
45. Leventhal GE, Kouyos R, Stadler T, et al. Inferring Epidemic Contact Structure from Phylogenetic Trees. *PLoS Comput Biol* **2012**; 8(3): 10.
46. Prem K, Cook AR, Jit M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput Biol* **2017**; 13(9): e1005697.

**Table 1.** Characteristics of patients included in the analyses.

Number of MSM included	2539
Age at registration in the SHCS (years, median[IQR])	36 [30-43]
Year of registration in the SHCS (years, median[IQR])	2007 [2001-2011]
ART ever started (N [%])	2484 [98%]
Use of recreational drugs during follow-up* (N [%])	1105 [44%]
Origin	
European	2303 [91%]
Education**	
Low	266 [11%]
Intermediate	1095 [43%]
High	1178 [46%]

\* Includes injected and not injected heroin,  
cocaine and “others” but excludes cannabis;

\*\*Low: Mandatory school or lower; High:

University

MSM: men who have sex with men; SHCS: Swiss

HIV Cohort Study; ART: antiretroviral therapy.

## Figure legends

### Figure 1. Graphical representation of the evolution the nsCAI status in MSM (Panel A)

**and aggregated trend (Panel B).** Each horizontal line in panel A represents the trajectory of an individual. Color changes along the x-axis reflect transitions between states (no nsCAI, nsCAI and not under follow-up). Variability along the y-axis reflects heterogeneity between patients.

White: not under follow-up, green: no nsCAI; blue: nsCAI; MSM: men who have sex with men; nsCAI: condomless anal intercourse with non-steady partners.

### Figure 2. Behavioural clusters of MSM and nsCAI trends by cluster.

The black vertical arrows indicate i) the advent of the «Swiss statement»/concept U(undetectable) = U(untransmissible) [Vernazza, Bull Med Suisse. 2008; Cohen, N Engl J Med 2016 ] and ii) scientific releases of landmark studies reporting successful Pre-exposure prophylaxis (PrEP). nsCAI: condomless anal intercourse with non-steady partners.

### Figure 3. Exemplary nsCAI trajectories and cluster membership.

Each horizontal line represents the trajectory of an individual. White: not under follow-up, green: no nsCAI; blue: nsCAI; Featured trajectories were selected at random. Hint: The estimated trajectories indicate that: Patient C1\_P1 did not engage in nsCAI until 2011, from when he continued without interruptions. Patient C4\_P2 did not engage in nsCAI until 2002, interrupted it in 2005 and engaged in nsCAI again in 2017.

### Figure 4. Receiver operator characteristic curves.

Variables tested for discriminatory power were education, origin, year of registration in the cohort and age.



BDT: Boosted decision trees; Likelihood: Maximum likelihood; kNN: k-nearest neighbours.

ROC curve interpretation hint: Larger areas between the ROC curve and the diagonal (grey) line indicate discriminatory power of the assessed variables.

**Figure 5.** Outcomes of the Toy Monte Carlo permutations tests comparing the area under the ROC (ROC AUC) displayed in Figure 4 (red) with that of 3000 runs with clusters labels assigned at random (blue).

**Figure 6. Configuration and correlations of sociodemographic variables across clusters.** The diagonal i.e, panels A, F, K and P) shows smoothed histograms, the rest of the panels, correlations between these variables. Age: in years in 2018. Origin labels: 0 (other), 1 (white), 2 (black), 3 (hispano-american), 4 (asian), 5 (unknown). Education labels: 0 (Low), 1(Intermediate), 2 (High).

Figure 1

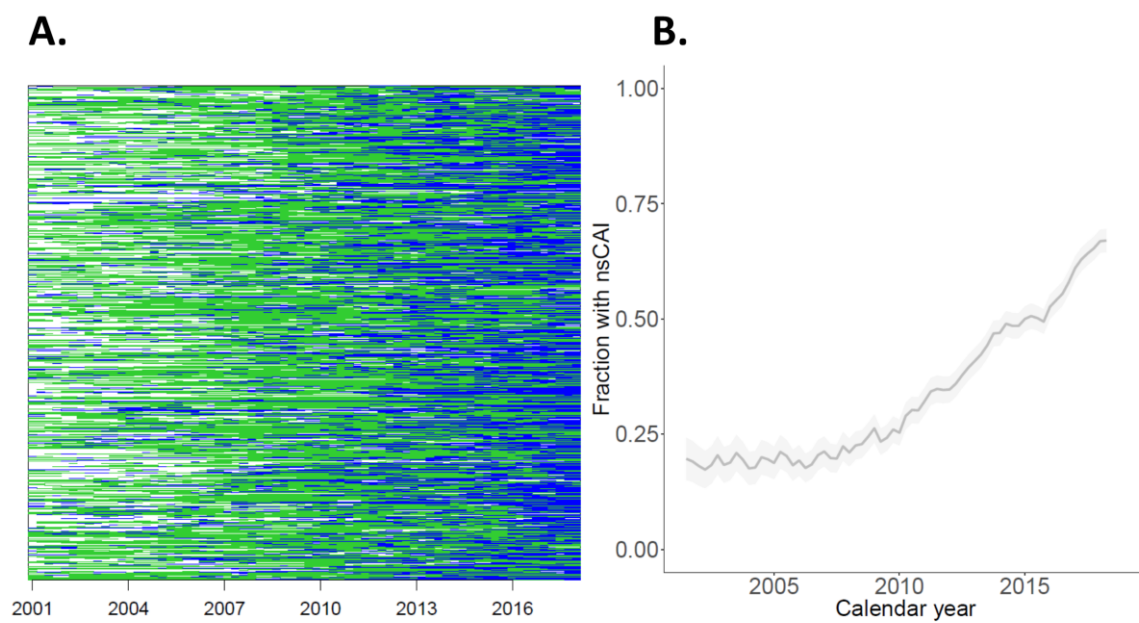
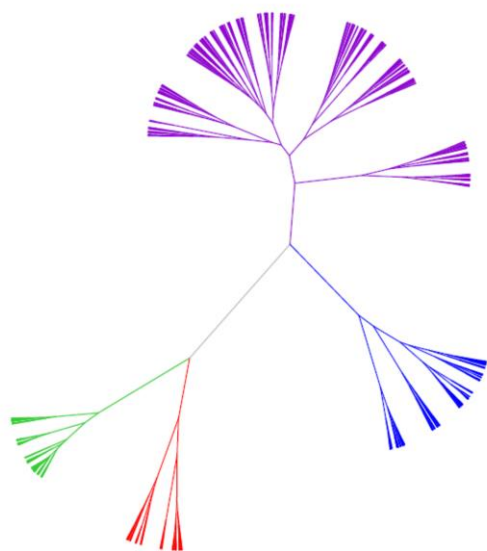


Figure 2

**A.**



**B.**

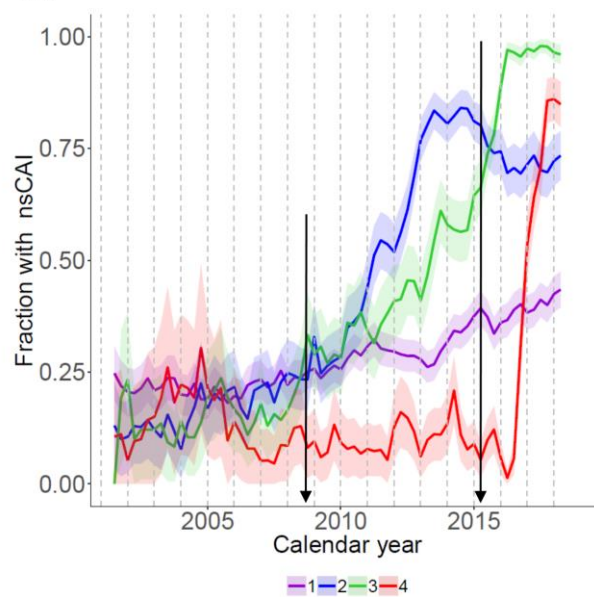


Figure 3

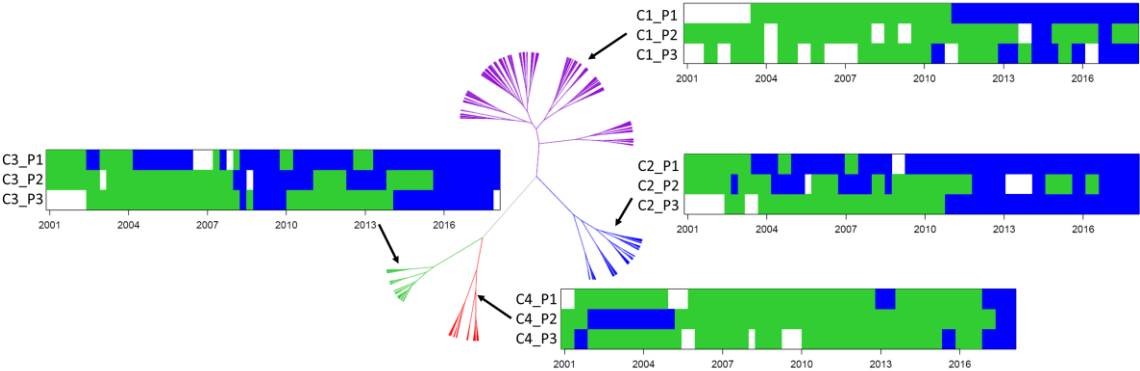


Figure 4

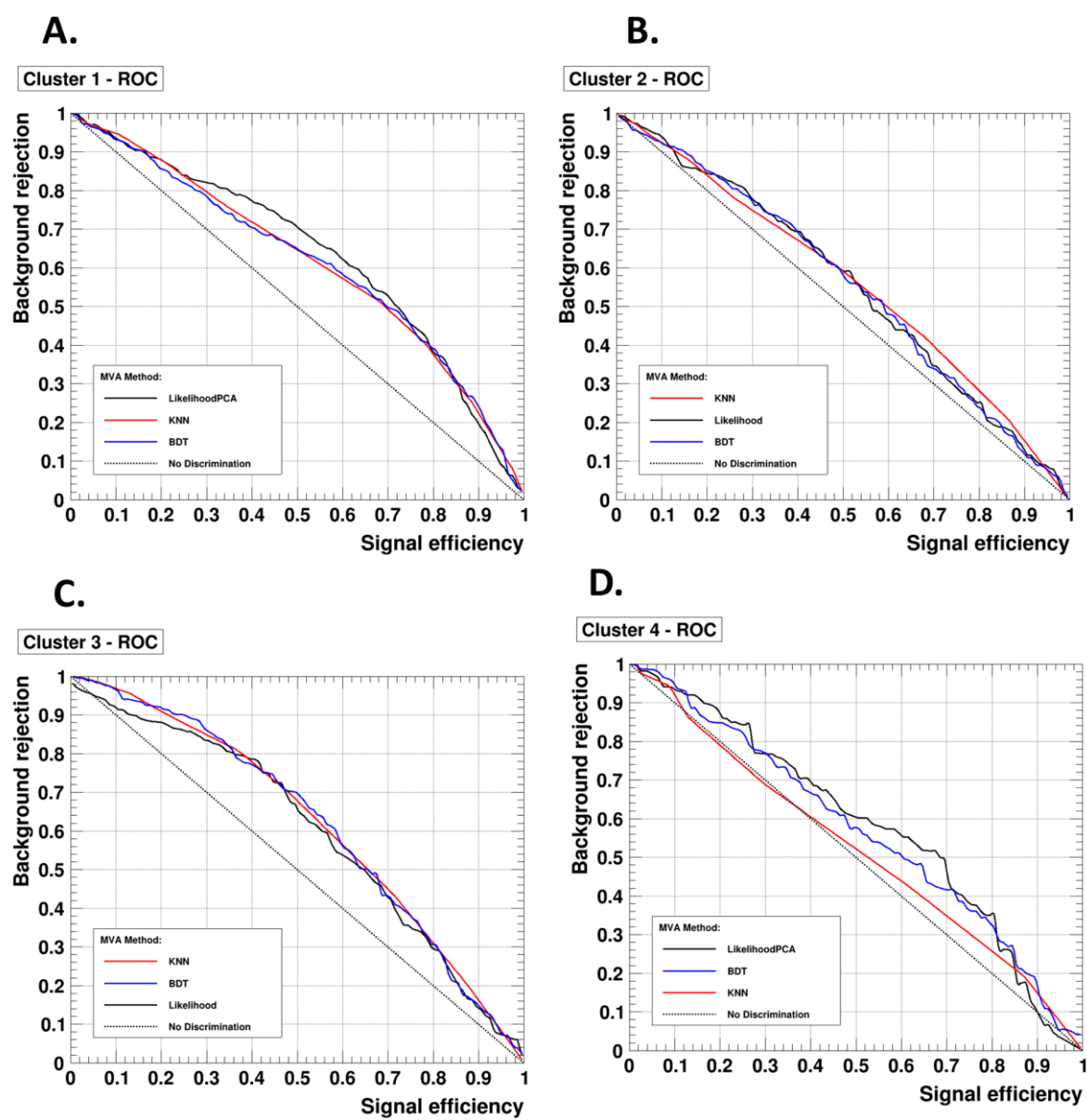


Figure 5

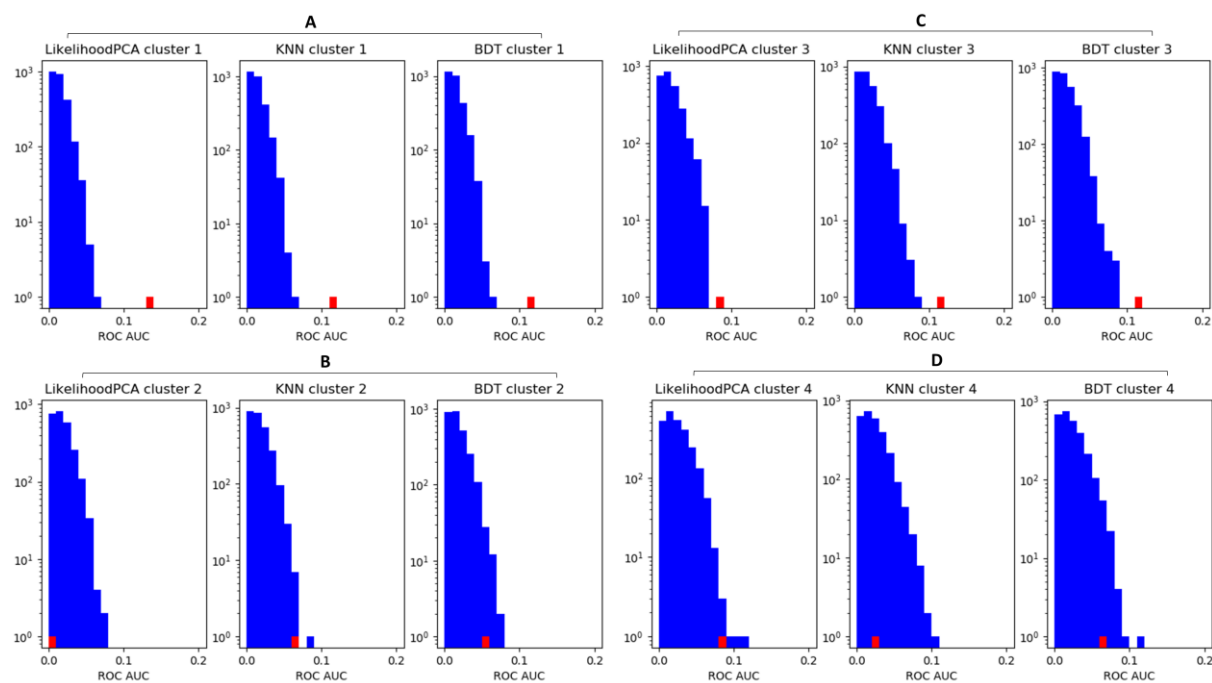


Figure 6

